DOCUMENT RESUME

ED 394 992 TM 024 773

AUTHOR Stroud, Robert E.
TITLE Correlations between the Metropolitan Achievement Tests, Seventh Edition, and the Ohio Ninth-Grade Proficiency Tests.
PUB DATE [95]
NOTE 12p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Comparative Analysis; *Correlation; Grade 8; Grade 9; High School Students; Language Arts; Mathematics Tests; Reading Tests; Secondary Education; Secondary School Students; Test Results; Test Use; *Test Validity; Writing Tests
IDENTIFIERS High Stakes Tests; *Metropolitan Achievement Tests; *Ninth Grade Proficiency Test

ABSTRACT

This study examines correlations between the Metropolitan Achievement Tests, seventh edition (MAT-7), and analogous portions of the 1995 Ohio Ninth-Grade Proficiency Tests. The MAT-7 scores of 156 eighth-grade students who had completed both test batteries were paired with complementary sections of the Proficiency test. Correlations between the tests were 0.52 for reading, 0.63 for mathematics, 0.25 for language/writing, and 0.58 for social studies and citizenship. All were considered significant. It is argued that even relatively high correlations between a test in question and an established test like the MAT-7 do not constitute validity. Validity comes only with the accumulation of results from different sources. Implications of comparing a high-stakes test like the Ohio Proficiency Test with a test administered in low-stakes circumstances are discussed. (Contains one table and nine references.) (Author/SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 394 992

Running Head: CORRELATIONS BETWEEN THE MAT-7 AND PROFICIENCY TESTS

# Correlations Between the Metropolitan Achievement Tests, Seventh Edition, and the Ohio Ninth-Grade Proficiency Tests

Robert E. Stroud

Bowling Green State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBERT E. STROUD

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

TM024773

BEST COPY AVAILABLE

Abstract

This study examines correlations between the Metropolitan Achievement Tests, seventh edition (MAT-7), and analogous portions of the 1995 Ohio Ninth-Grade Proficiency Tests. The MAT-7 scores of 156 eighth-grade students who had completed both test batteries were paired with complimentary sections of the Proficiency. Correlations between both tests were .52 for reading, .63 for math, .25 for language/writing, and .58 for social studies and citizenship. All were considered significant.

# Correlations Between the Metropolitan Achievement Tests, Seventh Edition, and the Ohio Ninth-Grade Proficiency Tests

Despite their detractors, aptitude and achievement tests remain a dominant tool for assessing the progress of not only students, but also schools and curricula. For years, state and local authorities have pressured lawmakers and administrators to adopt "rigorous" minimum competency standards while 1990 saw a National Assessment of Educational Progress panel meeting to set national standards of achievement (Rothman, 1990).

Passage and implementation of such tests is inevitably met with skepticism by those that question the validity of such measures. While it is true that no test can measure every facet of a child's academic development, it is often hoped that by comparing results of newly developed measures against established standards a test (or test battery) will be found both valid and reliable. Assessment tests such as the recently developed Basic Academic Skills Samples (BASS) have been measured not only against the MAT series, but the Gates-MacGinitie Reading Tests and Wide Range Achievement Test Revised (WRAT-R) as well (Jenkins & Jewell, 1992). In a similar instance, Dunn, McGhee, and Bryant (1992) found strong correlations (0.68 to 0.78) between the newly developed Detroit Tests of Learning Aptitude-Primary (DTLA-P:2) and the Woodcock-Johnson Psycho-Educational Battery (WJPB), a test series which had displayed established validity (Dunn, et al., 1992).

This strategy is useful in examining differences in scores as well as similarities. Han and Hoover (1994) utilized scores of the

Iowa Test of Basic Skills (ITBS), Iowa Test of Educational Development (ITED), and the Tests of Achievement and Proficiency (TAP) over a twenty-nine year period to assess gender differences in scores over time (Han, Hoover, 1994).

Consistency among tests enables their use in predictive studies as well, as was evidenced by Alex L. Chew and John D. Morris, whose 1989 studies involving the Metropolitan Readiness Test (MRT) and the Lollipop Test, indicated that both were reliable predictors of student achievement in successive years (Chew & Morris, 1989).

It should not be inferred from the previous text that either single, or even double correlations will lead to complete confidence. Indeed, Anastasi (1988) cautioned that test validity comes only with the accumulation of results from different sources (referenced in McGhee et al, 1992). The fact that a particular group of eighth-graders were given both the MAT-7 and the Proficiency Tests within the same school year does not provide an all-inclusive opportunity to assess the merits of the Proficiency. It does provide an opportunity to answer the question "are there any correlations between the two test scores?"

## METHODS

### Participants

The sample comprised 156 eighth-grade students from a single suburban junior high school in the Midwest. The sample consisted of 69 males and 87 females. Though students with Special Education classification participated in both test batteries, their

Proficiency scores were not reported and subsequently were not included in the 156 student sample.

Measures

The Metropolitan Achievement Tests is a battery consisting of 7 tests. The reading test, with a possible raw score of 85, consists of two sections, vocabulary and reading comprehension, with possible scores of 30 and 55 respectively. The mathematics test, with a possible raw score of 78, also contains two sections, concepts and problem solving (54 points) and procedures (24). The language test (54 points total) is comprised of three sections, prewriting (15), composing (15), and editing (24). While the MAT-7 contains four additional tests, only social studies (40) was used for this study, as it was the only test analogous to the citizenship portion of the Proficiency Tests.

The Ohio Ninth-Grade Proficiency Tests are, likewise, a battery of tests, each broken into specific sub-tests. Student scores are reported as either passing, which is not accompanied by a score, and failing which is accompanied by a score. For the purposes of this study, Proficiency results were scored as either passing (score of 1), or failing (score of 0). The breakdown for the Proficiency Tests are as follows:

Writing consists of content and organization, language, and writing conventions. Reading consists of fiction and non-fiction, each of which is broken down further into construes meaning and extends meaning. Everyday functional completes the reading test. Mathematics consists of measurement, arithmetic, geometry, data

analysis, and algebra. Finally, citizenship is broken down into geography, citizen knowledge, government, economics, law, and history.

Procedures

The MAT-7 was administered in the Fall of the students' eighth-grade year and the Proficiency Tests were taken the following Spring. Both tests were given in a group setting. The MAT-7 uses a multiple-choice format while the Proficiency employs both multiple-choice and written responses by students. Scores were gathered in the Summer of 1995 at which time, student scores on both tests were paired. Pairing of the tests was as follows:

MAT-7 reading/Proficiency reading, MAT-7 math/Proficiency math, MAT-7 language/Proficiency writing, and MAT-7 social studies/ Proficiency citizenship.

## Results

The correlations for the MAT-7/Proficiency pairings are shown in table 1 (alpha level = .05, two tails). For brevity, MAT-7 reading is abbreviated MREAD, Proficiency reading is abbreviated PREAD, and so on. While the original correlation matrix included correlations between all tests, only the paired tests correlations are shown.

(put table 1 here)

In all pairings, correlations were significant with $r = +.25$, $n = 156$, $p < .01$, two tails.

## Discussion

To restate an earlier point, even relatively high correlations

between a test in question and an established test (like the MAT) do not constitute validity. The MAT series are widely used and are expected to correlate significantly with other achievement tests, but as Jenkins and Jewell (1992) noted upon review of the MAT, the technical manual "offers no specific data to support this claim" (p. 278). Some critics have gone beyond merely questioning validity, and implied that tests can actually do psychological harm. Edward Burns, in his book The Development, Use and Abuse of Educational Tests (1979), hypothesizes that many students become traumatized by repeated exposure to tests that they fully expect to do poorly on. He feels that educators place too much value upon these scores, and even dedicates his book to "my children, and to all children, who have been baptized into the mystique of educational testing" (see dedication).

Test content may be found inappropriate on moral grounds as well as statistical. The California Learning Assessment System (CLAS), originally administered in 1993, has met with resistance from conservative and/or religious parents who found the material in several of the test questions objectionable (Colvin, 1995). The CLAS was also held suspect because only a portion of the results were scored and reported. This was an effort to save money, but ended up tainting the test's reputation instead.

Even when tests are administered and scored correctly, there remains the question of how to report the results. A child's score is often reported as a percentile rank, something parents often seem comfortable with. However, percentile rank is a concept not

understood by all, and the same may be said of other reporting methods as well (i.e. raw score, stanine, scaled scores, etc.)

A final point, and one of considerable importance to this study, is the question of student motivation. During administration of the MAT-7, the students in this sample were encouraged to do their best but were also aware that their scores would not effect their graduation. The same cannot be said for the Proficiency, as the students were made aware that they must eventually pass all four tests. The school in which this testing took place offered additional incentives for passing as well, including waiver of final exams and early release from school. Similar questions were raised in Georgia after 1993 results of the ITBS showed little student progress, particularly in the eleventh-grade where students knew their ITBS scores would not be recorded on their permanent record (White, 1993).

Despite the aforementioned difficulties, educational testing does have its uses. Baker (1982) acknowledges the virtues of testing in the following statement:

> Tests are important because they fulfill three general functions. First, they allow for some aspects of education to become public . . . . (Second, they) are assumed to permit insight into the quality of educational efforts. This insight relates closely to accountability . . . . (Thirdly) people have assumed that having tests assures that schools have standards of quality . . . (p. 1).

Baker's points seem to suggest that the true calling of educational

testing is to sate public desire for information, and this insight into public relations cannot be overlooked. Perhaps, though, for the professional educator or researcher, true insight into student (and school) performance comes from review of not only a battery of tests, but from GPA, student report, and indeed every avenue of information available as well.

References

Baker, E. L. (1982). Issues in achievement Testing. Los Angeles, California: Center For The Study Of Evaluation.

Burns, E. (1979). The development, use, and abuse of educational testing. Springfield, Illinois: Charles C. Thomas.

Chew, A. L., & Morris, J. D. (1989). Predicting later academic achievement from kindergarten scores on the metropolitan readiness tests and the lollipop test. Educational and Psychological Measurement, 1989, v.49, 461-465.

Colvin, R. L. (1995, April 5). State test finds students lagging. Los Angeles Times, pp. A1, A22.

Dunn, C., McGhee, R., Bryant, B. R. (1992). A validation study of the detroit tests of learning aptitude-primary second edition. Diagnostique, 1992, 17, 266-272.

Han, L., Hoover, H. D. Gender differences in achievement test scores (1994) [ERIC]. New Orleans, LA. National Council on Measurement in Education, EDRS Price (producer and distributor respectively).

Jenkins, J. R., Jewell, M. (1992). An examination of the concurrent validity of the basic academic skills samples (BASS). diagnostique, 1992, 17, 273-288.

Rothman, R. (1990). Despite doubts, NAEP panel meets to set first national standards for achievement. Education Week 10, 1, 1, 42.

White, B. (1993, June 3). Higher test scores elude state's pupils. Atlanta Constitution, p. 6, col. 1.

Table 1.

| | PREAD | PMATH | PWRIT | PCITI |
|---|---|---|---|---|
| MREAD | r = + .52 | | | |
| MMATH | | r = + .63 | | |
| MLANG | | | r = + .25 | |
| MSOC | | | | r = + .58 |

n = 156

p < .01, two tails